2017

Reliable Estimation of Minimum Embedding Dimension Through Statistical Analysis of Nearest Neighbors

David Chelidze
University of Rhode Island, chelidze@uri.edu

Citation/Publisher Attribution

Chelidze, D. (2017). Reliable Estimation of Minimum Embedding Dimension Through Statistical Analysis of Nearest Neighbors. *J. Comput. Nonlinear Dynam., 12*(5), 051024. doi: 10.1115/1.4036814
Available at: http://dx.doi.org/10.1115/1.4036814

Reliable Estimation of Minimum Embedding Dimension Through Statistical Analysis of Nearest Neighbors

Reliable Estimation of Minimum Embedding Dimension Through Statistical Analysis of Nearest Neighbors

David Chelidze

Professor, Member of ASME

Nonlinear Dynamics Laboratory

Department of Mechanical, Industrial and Systems Engineering

University of Rhode Island, Kingston, RI 02881, USA

Email: chelidze@uri.edu

ABSTRACT

False nearest neighbors (FNN) is one of the essential methods used in estimating the minimally sufficient embedding dimension in delay-coordinate embedding of deterministic time series. Its use for stochastic and noisy deterministic time series is problematic and erroneously indicates a finite embedding dimension. Various modifications to the original method have been proposed to mitigate this problem, but those are still not reliable for noisy time series. Here, nearest-neighbor statistics are studied for uncorrelated random time series and contrasted with the corresponding deterministic and stochastic statistics. New composite FNN metrics are constructed and their performance is evaluated for deterministic, stochastic, and random time series. In addition, noise-contaminated deterministic data analysis shows that these composite FNN metrics are robust to noise. All FNN results are also contrasted with surrogate data analysis to show their robustness. The new metrics clearly identify random time series as not having a finite embedding dimension and provide information about the deterministic part of stochastic processes. These metrics can also be used to differentiate between chaotic and random time series.

1 Delay-Coordinate Embedding

The embedding theorem [1] has paved the way to phase-space-based nonlinear time series analysis. It requires only two parameters: the embedding dimension and the delay time [2]. Given a deterministic scalar time series $\{x_i\}_{i=1}^{N} \in \mathbb{R}$ sampled uniformly with t_s sampling time, reconstructed phase-space vectors, $\{\mathbf{x}_i\}_{i=1}^{N-(d-1)\tau} \in \mathbb{R}^d$, are usually given by $\mathbf{x}_i = [x_i, x_{i+\tau}, \ldots, x_{i+(d-1)\tau}]^T$, where d is the embedding dimension, τt_s is the actual delay time, and τ is a positive integer or digital delay time. For the appropriate choice of delay parameters (d and τ), the resulting reconstruction is *diffeomorphically* equivalent to the original dynamical system's phase space [2]. There are other generalizations to delay-coordinate embedding with variable delays [3, 4], and they can yield lower-dimensional embeddings (e.g., in case of quasiperiodic signal [4]). Here, we do not consider these generalizations and focus on obtaining optimal constant delay embeddings.

The estimation of optimal delay time can be approached from a purely geometrical perspective [5] or by considering linear/nonlinear (auto)correlations in the time series [6, 7]. In some cases, selection of the optimal delay time and embedding dimension cannot be decoupled (e.g., when mutual information changes with the embedding dimension [8]). Some studies have advocated focusing on the so-called *embedding window* ($\tau_e = (d-1)\tau$) that describes the total time span covered by all coordinates in each reconstructed phase-space points [9–12]. Others use model-based parameter estimation [12–16] to estimate both optimal delay and embedding dimension simultaneously. More recently, the optimal short-term prediction metric (based on artificial neural network model) was used for estimating both appropriate delays and embedding dimension [17]. A similar idea, in conjunction with average mutual information for the delay estimation, was used in estimating embedding dimension in Ref. [18]. Some of these methods are applicable in certain scenarios, others have some deficiencies, as discussed in Refs. [8, 19, 20]. However, to focus on estimating minimum embedding dimension, we either assume that the delay is determined *a priori* or chosen conservatively so that a spatial decorrelation transform [21] can be used. This restriction allows for a compact comparison between different methodologies considered in this paper.

Here we focus on one of the popular methods used for estimating the minimally sufficient embedding dimension called *false nearest neighbors* (FNN) [22]. There are other methods that utilize convergence in the estimates of some nonlinear measures or metrics [23] but they are either more complex or cumbersome in the applications. The idea behind the FNN [22] method is based on the uniqueness property of a dynamical system's phase-space trajectory. The *nearest neighbor* (NN) of a point in a d-dimensional embedding is labeled false if the pair of points are close only due to the projection from a higher-dimensional (e.g., $(d+1)$-dimensional) phase space. Thus, the FNN will separate if the data is

embedded into a $(d+1)$-dimensional space, while the *true* NNs will remain close.If one is able to detect all the FNN, then the minimally sufficient embedding dimension can be identified as the least dimension needed to achieve zero fraction of the FNN.

There are three distinct variations of the FNN algorithm. We label them as: the *original* or *Kennel* [22], *improved* [24], and *reliable* [21] algorithms. Variations of them are frequently used in very diverse fields such as stock market price forecasting [25], astrophysics [26], engineering [27], and biology [28]. We will discuss each of the algorithms and their properties with applications to deterministic, stochastic, and uncorrelated random time series. Particular attention will be given to the observed deficiencies in each of the algorithms and the proposed solutions will be discussed and tested. The effect of additive noise in deterministic time series on the performance of the algorithms will also be considered.

2 Data Sets Used in the Analysis

For the purely white random time series, we generate 6×10^4 point samples of normally and uniformly distributed random numbers. Both uncorrelated random data sets are normalized to have unit variance, and the arbitrarily chosen $\tau=5$ (the choice is arbitrary since there are no nonzero autocorrelations in these time series for $\tau\neq0$) is used in delay-coordinate embedding. In what follows, we will label these random data sets as *Normal* and *Uniform*, correspondingly.

For the stochastic data containing deterministic components, we use an autoregressive process of order two (a discretized noise-driven damped oscillator, just like in [24]):

$$x_{n+1}=(2-\omega^2-\rho)x_n+(\rho-1)x_{n-1}+\eta_n\,, \tag{1}$$

where η_n is white noise, $\omega=2\pi/20$ is the frequency, and a total of four 6×10^4-point time series are generated for each value of $\rho=[0.02,0.05,0.2,0.5]$. The first two values of ρ represent the small amount of stochasticity in mainly deterministic data, which can also be viewed as deterministic data contaminated by multiplicative noise. In contrast, the latter two values of ρ reflect dominant randomness. In what follows, each of these time series will be labeled AR(0.02), AR(0.05), AR(0.2), and AR(0.5), respectively. The delay $\tau=5$ is used for all four time series as dictated by the natural frequency of the oscillator, autocorrelation, and the average mutual information (see Fig. 1).

In addition to the white random and correlated stochastic time series, we also include data for two chaotic time series. The first is generated from the Lorenz equation [29]:

$$\dot{x}=-\frac{8}{3}x+yz,\;\; \dot{y}=-10\,(y-z)\;,\;\; \&\;\; \dot{z}=-xy+28y-z\,. \tag{2}$$

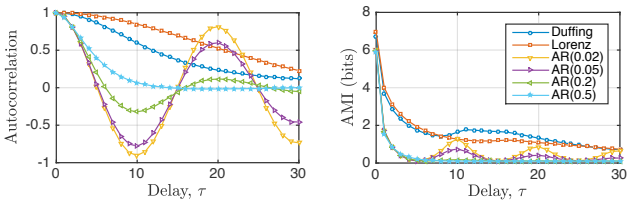

Fig. 1. Linear and nonlinear correlations in chaotic and correlated stochastic time series

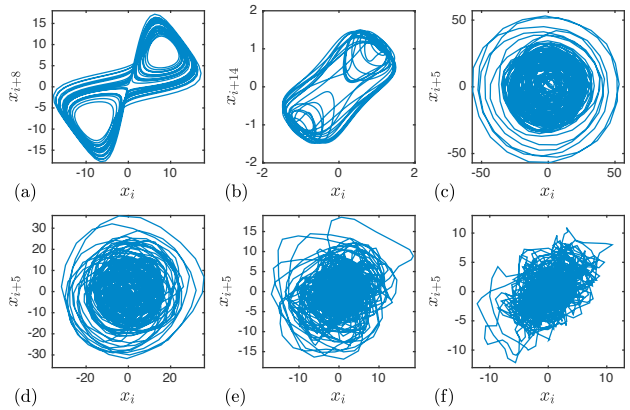

Fig. 2. Reconstructed phase portraits for Lorenz (a), Duffing (b), AR(0.02) (c), AR(0.05) (d), AR(0.2) (e), and AR(0.5) (f)

The Lorenz time series is sampled with a $t_s = 0.02$ time interval and a total of 6×10^4 points are recorded starting from the initial condition $(x_0, y_0, z_0) = (20, 5, -5)$. In the following analysis, we only use the y coordinate, with $\tau = 8$ delay as indicated by the mutual information (Fig. 1), and label the corresponding results as *Lorenz*.

The second chaotic time series is generated using the double-well Duffing equation [30]:

$$\ddot{x} + 0.2\dot{x} - x + x^3 = 0.33\cos t\,. \tag{3}$$

The total of 6×10^4 steady state response points are recorded using a $t_s = 0.1$ sampling time interval. In the analysis, only the x time series is used with $\tau = 14$ delay, also indicated by the mutual information (Fig. 1), and the results are labeled as *Duffing*. The reconstructed phase portraits of deterministic and correlated stochastic time series are shown in Fig. 2.

For the Lorentz and Duffing equations, using just scalar time series, one expects to need three- and four-dimensional embeddings, respectively, to eliminate all the FNN [19, 20, 31]. Strictly speaking, the embedding theorem [2] guarantees the embedding if the dimension is greater than twice the fractal dimension of an attractor (i.e., for $d = 5$ for both Lorenz and Duffing). The white noise time series are theoretically infinite dimensional. However, for our finite time series, all the false NNs will be eliminated if the embedding dimension is equal to the data size (i.e., we have only one embedded data point). The same is also true for the correlated stochastic time series in a strict sense. However, these correlated random

data all possess a deterministic damped oscillator backbone that is low-dimensional and is guaranteed to be embedded in three dimensions, while just two-dimensional embedding should generically work, too. Thus, one would expect the decrease in FNN due to the presence of a deterministic structure in the data, especially for large correlation times.

Additive noise effects are investigated by adding four different noise levels to the Lorenz and Duffing data. In particular, 5 %, 10 %, 20 % and 40 % noise (determined by the ratio of the standard deviation of noise σ_n over the standard deviation of signal σ_x) are added to the time series and the results are labeled accordingly. By the conventional definition of *signal to noise ratio* (SNR) as $10\log_{10}\sigma_x^2/\sigma_n^2$, these correspond to 26.02 dB, 20 dB, 13.98 dB, and 7.96 dB SNR, respectively.

All the data sets are successively embedded into $d = 1,\ldots,20$ dimensions. For each d-dimensional embedding we find the NN distance for each point, as well as the distance between the $(d+1)$-th components of the same pair of points. In addition, results are contrasted with a surrogate time series analysis similar to the one used in [21]. Instead of going through an exercise of frequency and time domain randomization and normalization procedures [32], they have used a much simpler strategy. For each d-dimensional NN pair, the $(d+1)$-dimensional coordinate distance is estimated by randomly selecting one of the coordinates from all the data. This is justifiable, if all the linear correlations are absent from the components of the reconstructed phase space (the needed coordinate transformation is discussed in Section 3.3.1), where the noise would appear to be effectively white. Other algorithmic details are described in the following sections.

3 Overview of False NN Algorithms

3.1 The Original Algorithms

The original FNN method was proposed by Kennel *et al.* in 1992 [22]. Working in each d-dimensional reconstructed phase space, for each point $\mathbf{x}_i$ its NN $\mathbf{x}_{j(i)}$ is *identified as false* if:

$$\left|x_{i+d\tau} - x_{j(i)+d\tau}\right| > r\left\|\mathbf{x}_i - \mathbf{x}_{j(i)}\right\|, \tag{4}$$

where r is an *a priori* fixed threshold value (typically chosen to be 10 in Ref. [22]).

To simplify notation and streamline discussion, we define the following variables:

$$\delta_i \triangleq \left|x_{i+d\tau} - x_{j(i)+d\tau}\right|, \quad \text{and} \quad \varepsilon_i \triangleq \left\|\mathbf{x}_i - \mathbf{x}_{j(i)}\right\|. \tag{5}$$

In what follows, we will also imply the functional dependence of index j on the index i (i.e., $\mathbf{x}_j$ is always the NN to $\mathbf{x}_i$

using the appropriate NN metric). Then, the probability that any point $\mathbf{x}_i$ and its NN $\mathbf{x}_j$ are FNN can be written as:

$$f_{nn}(d;r) \triangleq \text{Prob}\left\{\delta_i > r\varepsilon_i\right\}, \tag{6}$$

and we refer to it as *original fraction*. Then, the idea is that the smallest dimension d for which $f_{nn}(d;r)$ reaches zero is the minimally *sufficient* embedding dimension. The original fraction provides misleading results for the white random and correlated stochastic time series for which it erroneously shows low embedding dimensions and the surrogate analysis does not provide the differentiation [21]. In fact, results for uncorrelated random time series and for large amplitude autoregresive processes are virtually identical and indicate the existence of a low-dimensional attractor. This is a problem for noisy deterministic time series, since we cannot definitively attribute zero FNN to the actual existence of a low-dimensional attractor.

Hegger and Kantz [24] were first to clearly identify the shortcomings of this definition when noisy deterministic data were considered. For uniformly distributed white noise, the original fraction $f_{nn}(d;r) = 1 - 2\langle\varepsilon_i\rangle\, r + \langle\varepsilon_i^2\rangle\, r^2$, where $\varepsilon_i = \varepsilon_i(d)$ is the distance between the NNs in d dimensions. For uncorrelated noise time series of size N, $\langle\varepsilon_i\rangle \simeq N^{-1/d}$. Therefore, this FNN fraction approaches zero when $\langle\varepsilon_i\rangle \simeq r^{-1}$ or when $N \simeq r^d$. To always differentiate noise (which is inherently infinite-dimensional) from deterministic data (finite-dimensional), one always needs $N \geq r^d$ to get the nonzero original fraction for uncorrelated noise time series. Unfortunately, this is not always feasible for experimental data.

To address this problem, in [22], the NNs were also counted as false if the distance in $(d+1)$-dimensions was greater than the a times σ_x of the data (e.g., $a = 2$ in [22]). Thus, the FNN condition changed to:

$$\hat{f}_{nn}(d;r) \triangleq \text{Prob}\left\{(\delta_i > r\varepsilon_i) \vee \left(\delta_i^2 + \varepsilon_i^2 > a^2\sigma_x^2\right)\right\}, \tag{7}$$

which we call the *Kennel fraction*.

Figure 10 shows the results of this algorithm applied to our data sets. This solution seems to improve performance for the white random time series by increasing the FNN fraction for higher embedding dimensions. However, it exhibits a dip in the FNN fraction at lower dimensions (i.e., it underestimates the FNN fraction for large r and low d). This dip is also present for the correlated stochastic time series especially for ones with the larger stochastic components. However, the surrogates also show a similar dip and diverge from actual data after the minimum is reached. Even with this divergence the correlated stochastic data is hard to interpret and this solution is inadequate [21]. This will also be problematic for noisy deterministic time series, since this dip cannot be uniquely attributed to the existence of a low-dimensional attractor. In addition, [22] points out that this solution may introduce false neighbors in large d for an insufficient amount of deterministic data.

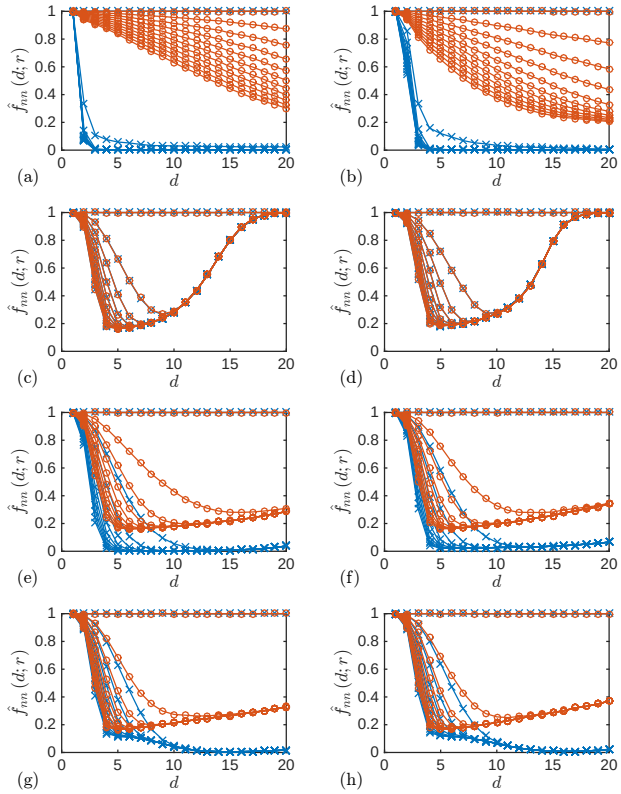

Fig. 3. Kennel fraction of FNN using Eq. (7) for Lorenz (a), Duffing (b), Normal (c), Uniform (d), AR(0.02) (e), AR(0.05) (f), AR(0.2) (g), and AR(0.5) (h). The gradual decrease in the FNN fraction is precipitated by incrementing r from zero to 20 by two. Lines with circles reflect the corresponding surrogate data results.

3.2 The Improved Algorithm

In [24], it is supposed that if the distance between the NNs is larger than σ_x, these points cannot be considered true NNs and should be discarded in computation. In this approach, the NNs for which the d-dimensional distances are greater or equal to $1/r$ times σ_x are disregarded and the original fraction is used for FNN. Unfortunately, for large values of d and r, this causes the number of available points to decrease drastically to zero for both white random and correlated stochastic data sets. The current algorithm for false NNs, as described in TISEAN package [19, 33] has the following form:

$$\tilde{f}_{nn}(d;r) \triangleq \frac{\sum_{i=1}^{N-(d-1)\tau} \Theta(\delta_i - r\varepsilon_i)\,\Theta(\sigma_x - r\varepsilon_i)}{\sum_{i=1}^{N-(d-1)\tau} \Theta(\sigma_x - r\varepsilon_i)}, \tag{8}$$

where Θ is the Heavyside function and $\tilde{f}_{nn}$ is referred to as the *TISEAN fraction*. They claim that this solution for white random data leads to 50–60 % FNN for large N, independent of choice of d and r. However, in our white random and dominantly stochastic test cases, the number of points suitable for averaging drops to zero precipitously for large d values. In addition, it indicates low-dimensional embedding for stochastic or random data and does not provide any differentiation

with the surrogate analysis.

3.3 The Reliable Algorithm

In [21], some of the ideas in [24] were further expanded and a new FNN algorithm was proposed in conjunction with new and interesting *false nearest strands* (FNS) idea. These new FNN and FNS methods were shown to be usable for noisy or stochastic time series. In [21] several important improvements to the FNN algorithm were proposed: (1) only NNs that are temporarily uncorrelated are considered to focus exclusively on spatial proximity; (2) instead of FNN, FNS were advocated to make sure only true NNs were considered in the calculations even for noisy data; and (3) a spatial decorrelation transform was introduced to deal with linear correlation bias due to small delays.

3.3.1 Spatial Decorrelation

The spatial decorrelation step is aimed at negating the effect of the too-small delay used in the reconstruction, which makes the FNN fraction small irrespective of the actual dynamics. In addition, this step makes the method insensitive to the appropriate choice of the time delay, unless it is chosen to be unreasonably large.

The basic idea of the decorrelation is to counteract the bunching of all reconstructed points near a thin tube along the hyperdiagonal of a phase space. Given the original $(d+1)$-dimensional reconstructed phase-space trajectory $Y \in \mathbb{R}^{(d+1)\times[N-\tau(d+1)]}$, where the i-th column is given by the i-th delay vector $\mathbf{x}_i \in \mathbb{R}^{d+1}$, Y is first transformed into its proper orthogonal space $\tilde{Y}$ using *singular value decomposition* [34]:

$$Y = UCV^T, \quad \Rightarrow \quad \tilde{Y} = U^T Y\,, \tag{9}$$

Then $\tilde{Y}$ is rotated in the phase space so that the first d-coordinates are *only* a function of the first d coordinates of the original phase space Y. To accomplish this we take the last column of $\tilde{Y}$, $\mathbf{u} \in \mathbb{R}^{d+1}$, and then the needed orthogonal transformation is given by the following Householder matrix [34]:

$$Q = H\left(\mathbf{u} - |\mathbf{u}|\,\mathbf{e}_{d+1}\right), \tag{10}$$

where $\mathbf{e}_{d+1}$ is the unit vector in the $(d+1)$ direction. Then the decorrelated phase space is given by:

$$\bar{Y} = Q\tilde{Y}\,. \tag{11}$$

In our test samples, the choice of delay time is close to optimal and the use of decorrelation does not provide any significant benefits, but in practice it can be a differentiating factor. Thus, we have included this step in the results presented from now on.

3.3.2 Temporal Decorrelation

As opposed to the TISEAN fraction, Ref. [21] advocates to look at just the distance between the $(d+1)$-th coordinates of the NNs. If we only consider the FNN- and *not* the FNS-based metric we will get:

$$\bar{f}_{nn}(d;s) \triangleq \text{Prob}\left\{\delta_i > s\,\sigma_x\right\}, \tag{12}$$

where, to identify true NNs for each point $\mathbf{x}_i$, its NN $\mathbf{x}_j$ is determined such that:

$$j = \underset{j}{\text{argmin}} \left\|\mathbf{x}_i - \mathbf{x}_j\right\|, \text{ subject to } |i-j| > w, \tag{13}$$

where w is a Theiler window [35, 36] used to remove temporal correlations. We call the metric defined by Eq. (12) the *reliable fraction*. Ref. [21] discusses the appropriate choice of the Theiler window and advocates to use two or three times the first minimum of average mutual information as a good rule of thumb. Another alternative is to match the w time scale to the point at which the average mutual information reaches its asymptotic zero value.

The results for Eqs. (12) and (13) using a Theiler window of four times the first minimum of the average mutual information are shown in Fig. 4 (left plots). For the deterministic time series, we observe an initial underestimation of the FNN fraction with the increase in s value. The fractions reach zero for an appropriate embedding dimension and then show gradual increase for small s values with the increase in d values. This gradual increase is more pronounced for the Duffing data and is absent for larger s values. For the white random data sets the FNN fraction behaves as expected, showing linear decrease as s is increased, while remaining constant for all values of d. For the correlated stochastic data this metric is more interesting and shows trends intermediate between the trends for the purely white random and deterministic data. Specifically, it can clearly be used with large s values to differentiate small stochastic component from the larger deterministic data—plots (e,f). The results for the surrogates are clearly separated from the actual data, especially for larger s values, and remain constant for all d.

3.3.3 False Nearest Strands

In the same paper [21], the shortcomings of the previous FNN methods for the deterministic data contaminated by small random noise are also identified. In this case, the NNs identified through the analysis may be proximal not due to the deterministic dynamics or the projections, but due to noise. If the proportion of the misidentified NNs is large, then the algorithm will overestimate the corresponding FNN fraction. To identify only true (in terms of deterministic dynamics and projection only) NNs, they propose to consider *NN strands* (NNS). The following procedure is used to identify NNS:

1. Identify NN $\mathbf{x}_{j(i)}$ for each point $\mathbf{x}_i$ given by Eq. (13).
2. If there is any pair of points (for a minimal $k = k^\star$) from a set of preceding NNs $\{\mathbf{x}_{i-k}, \mathbf{x}_{j(i-k)}\}_{k=1}^{w}$ such that $j(i) = k^\star + j(i-k^\star)$, then $[\mathbf{x}_i, \mathbf{x}_{j(i)}]$ are assigned to the strand containing $[\mathbf{x}_{i-k^\star}, \mathbf{x}_{j(i-k^\star)}]$. Otherwise, $[\mathbf{x}_i, \mathbf{x}_{j(i)}]$ are assigned to a new pair of NN strands.
3. After examining all the data, N_s number of NNS are obtained, each containing a set of NN points S_k ($k = 1, \ldots, N_s$). Then the k-th strand is designated false if $\langle \delta_i \rangle_k > s\,\sigma_x$, where

$$\langle \delta_i \rangle_k \triangleq \langle \delta_i \rangle_{i \in S_k} = \frac{1}{\|S_k\|} \sum_{i \in S_k} \delta_i \,. \tag{14}$$

Alternatively we can define *Reliable FNS Fraction* as:

$$\bar{f}_{ns}(d;s) \triangleq \mathrm{Prob}\left\{ \langle \delta_i \rangle_k > s\,\sigma_x \right\} . \tag{15}$$

The results of the FNS calculations for our test time series are shown in Fig. 4 (right plots). The decrease in the FNS fraction with the increase in s is still present as reliable FNN. For the deterministic data, some gradual increase in this fraction with the increase in d for small values of s is also present. In addition, and in contrast with the reliable FNN fraction, this metric does not converge to zero as fast and indicates larger minimal embedding dimension for the deterministic data. The results for the white random data are still indistinguishable from the ones for the surrogate data. However, the white random data trends show a gradual increase in the FNS fraction not observed reliable FNN. For the correlated stochastic data sets, FNS trends are clearly delineated from the surrogates for the low level of stochasticity, and show results similar to the white random data for the larger stochastic components. To accurately estimate embedding dimension in deterministic cases, one needs to set some threshold value for the FNS fraction since it does not reach zero as fast as in the reliable FNN fraction. This task is made easier by the clear separation from the surrogates FNS fraction, which stays relatively level with the increase in d. For the larger s values, the Duffing FNS fraction also indicates lower

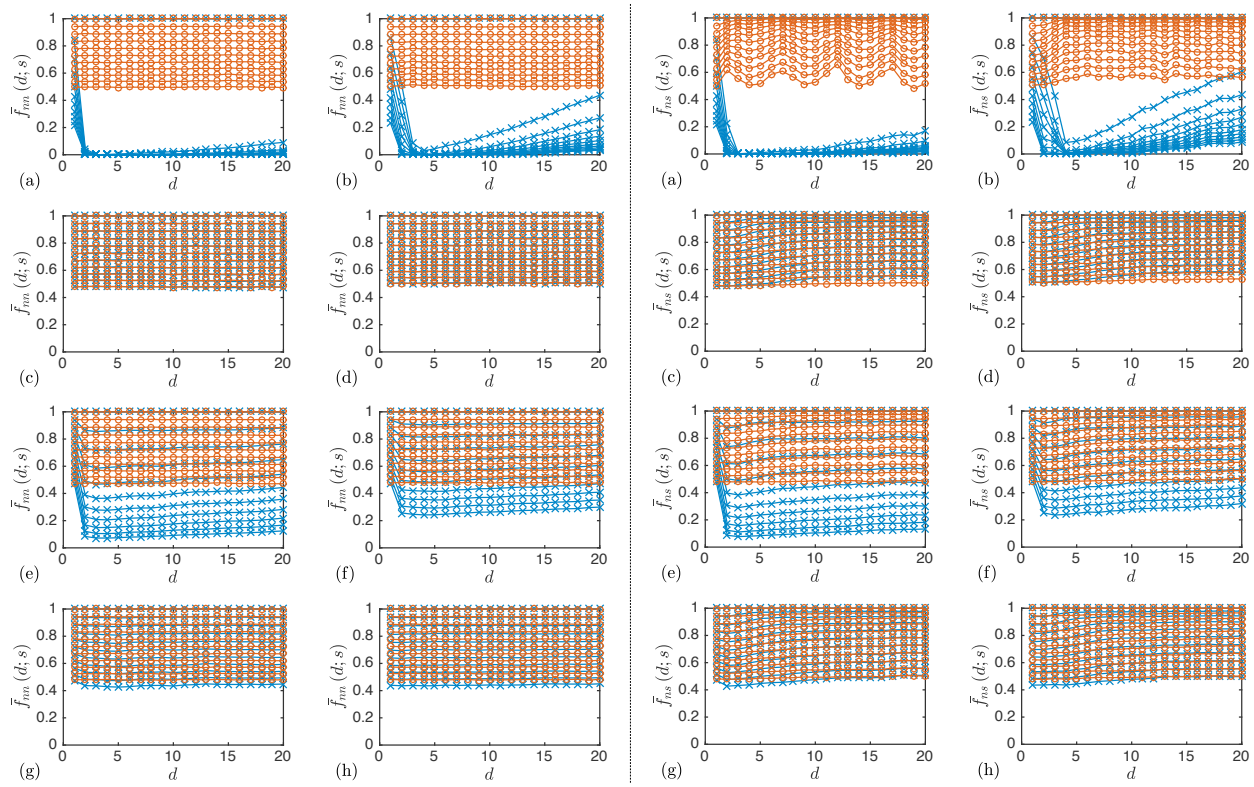

Fig. 4. Reliable FNN (left plots) and reliable FNS (right plots) fraction of Eq. (12) applied to Lorenz (a), Duffing (b), Normal (c), Uniform (d), AR(0.02) (e), AR(0.05) (f), AR(0.2) (g), and AR(0.5) (h). The gradual decrease in the FNN fraction is precipitated by incrementing s from zero to two by 0.1. Lines with circles reflect the corresponding surrogate data results.

than expected embedding dimension.

To see the advantage of these new metrics when dealing with noisy data, we apply them to Duffing data with additive Gaussian noise as shown in Fig. 5. For the reliable FNN algorithm, the new fraction floor rises with the noise level, but it is still usable for small values of s and low 5 % noise level Fig. 5(a). At the same time, the surrogate's FNN fraction stays constant with respect to d and asymptotes down with the increase in s to approximately 50 % fraction. The FNS data looks better for the low noise levels Fig. 5(b,d) and exhibits similar surrogate behavior. Therefore, it provides better separation between the surrogate and actual data FNS fractions compared to the FNN fractions. One might be able to estimate embedding dimension for the low noise levels in Fig. 5 if an appropriate threshold value is set. However, both reliable fractions cannot identify the embedding dimension for the larger noise levels.

In what follows, we want to further expand on some ideas in [21] and propose an FNN algorithm that is usable for noisy or stochastic time series with deterministic components. We begin with exploring the nearest-neighbor statistics for deterministic, white random, and correlated stochastic time series.

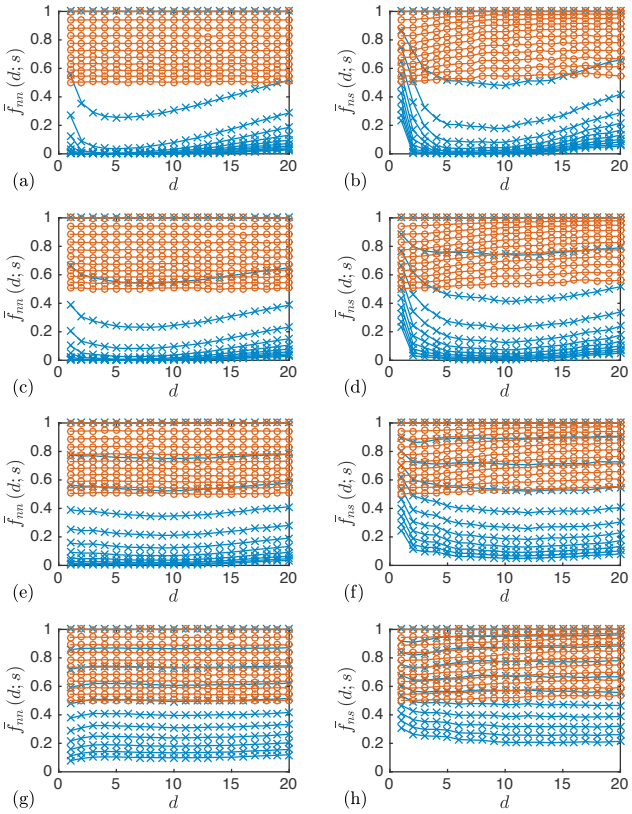

Fig. 5. Reliable FNN (first column: a,c,e,g) and FNS (second column: b,d,f,h) algorithm applied to Duffing data with 5 % (a,b), 10 % (c,d), 20 % (e,f), and 40 % (g,h) noise. The gradual decrease in the fractions is precipitated by incrementing s from zero to two by 0.1. Lines with circles reflect the corresponding surrogate data results.

4 Nearest-Neighbor Statistics

We hypothesize that the problem with the classical FNN lies with the definition of the fraction of FNN. To investigate this question, let us investigate NN statistics for our test time series. First, consider the expected average nearest-neighbor distance $\langle \varepsilon_i(d) \rangle$ for white random time series. Given a d-dimensional reconstruction of normally distributed random time series, the distance between any two points will be distributed according to a χ_d (d-dimensional chi distribution [37]). We also proposition that the expected value of the distance between the NN points will be proportional to the mean of the corresponding χ_d distribution and to the average density of points in d dimensions $(\sqrt{d}/N)^{1/d}$. This will result in the following approximation:

$$\langle \varepsilon_i(d) \rangle \simeq \sqrt{2} \frac{\Gamma((d+1)/2)}{\Gamma(d/2)} \left(\frac{\sqrt{d}}{N} \right)^{1/d}, \tag{16}$$

where Γ is the gamma function. Figure 6(a) shows the expected values of nearest-neighbor distances versus the embedding dimension for all eight test time series, without accounting for temporarily correlated NNs. The increase between the mean

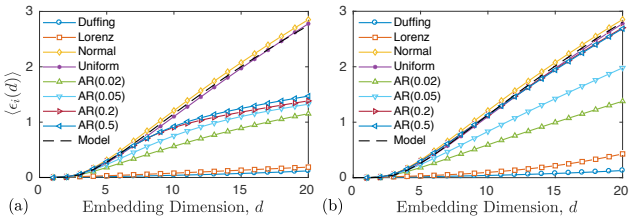

Fig. 6. Expected value of nearest-neighbor distances in d dimensions without (a) and with (b) temporarily decorrelated NNs

nearest-neighbor distance is more drastic for the white random data than the deterministic time series, and Eq. (16) follows this increase for the white random time series very closely.

The deterministic time series show more power-law type increase after the needed embedding dimension is reached ($d \sim 3$–4, in both case), and it has a considerably lower rate than for the random data. The nearest-neighbor distance for deterministic data is expected to grow according to local divergence rate $e^{\lambda_{\max} d \tau t_s}$, where $\lambda_{\max}$ is maximal Lyapunov exponent, after the minimum embedding dimension is reached. This can be expressed as:

$$\langle \varepsilon_i \rangle \approx \varepsilon \left(e^{\lambda_{\max} d \tau t_s} - 1 \right), \quad \forall d \geq m, \tag{17}$$

where ε is a constant reflecting the point density in the phase space. The parameters that follow the observed curves closely are: $\varepsilon = 0.0404 \pm 0.0008$ and $\lambda_{\max} \tau t_s = 0.0689 \pm 0.0008$ for the Lorenz data, and $\varepsilon = 0.5110 \pm 0.0742$ and $\lambda_{\max} \tau t_s = 0.0167 \pm 0.0020$ for the Duffing data. The exponential growth in Eq. (17) is only expected until it reaches the attractor size at which point it will saturate. Even if it is allowed to grow unimpeded it will only cross Eq. (16) somewhere $d \sim 60$. Thus, when dealing with low-dimensional deterministic systems (e.g., $d < 40$) their expected nearest-neighbor distance will stay well below the distance expected for the white random data in the same dimension.

The correlated stochastic time series show behavior somewhere in the middle of the white random and deterministic nearest-neighbor distance trends. They start close to the white random at low dimensions and then diverge to smaller distances for the larger embedding dimensions.

The same plot was regenerated for the distance between temporarily decorrelated NNs as shown in Fig. 6(b). The main observed differences are that the large stochastic component data now closely follows the white random distances, while the small stochastic component data shows the intermediate behavior. In addition, the deterministic data have more pronounced trends clearly showing differences in the divergence rates. Fitting the Eq. (17) to these curves gives the following parameters: $\varepsilon = 0.0333 \pm 0.0007$ and $\lambda_{\max} \tau t_s = 0.0790 \pm 0.0008$ for the Lorenz data, and $\varepsilon = 0.0425 \pm 0.0020$ and $\lambda_{\max} \tau t_s = 0.1212 \pm 0.0021$ for the Duffing data. Thus, we observe: (1) temporal correlations considerably stint the

divergence rates, especially for the Duffing data; and (2) divergence exponents after removing temporal correlations for Lorenz ($\lambda_{\max} \approx 0.5$) and Duffing ($\lambda_{\max} \approx 0.1$), while not being suitable for estimating maximal Lyapunov exponents, are close to their expected values [31] (experimentally estimated largest Lyapunov exponents from data are $\lambda_1 \approx 0.825$ for Lorenz and $\lambda_1 \approx 0.134$ for Duffing).

Now, let us consider the $(d+1)$-th coordinates of the NNs in d dimensions, which are still uncorrelated random numbers for the white random time series. Thus, the distance between these $(d+1)$-th coordinates for normally distributed variables ($x_i \sim \mathcal{N}[0,\sigma_x]$) follows a chi distribution [37][1], and the corresponding expected value can be expressed as:

$$\langle \delta_i \rangle = \mathbb{E}[\delta_i] = \frac{2\Gamma(1)}{\Gamma(1/2)}\sigma_x \approx 1.1284\,\sigma_x\,. \tag{18}$$

Now, it is easy to show that for the uniformly distributed variables (i.e., $x_i \sim \mathcal{U}[-a,b]$), this expected value becomes:

$$\langle \delta_i \rangle = \frac{2}{\sqrt{3}}\sigma_x \approx 1.1547\,\sigma_x\,. \tag{19}$$

We conclude that for some general uncorrelated random time series the average distance between the $(d+1)$-th coordinates is expected to be slightly greater than the corresponding σ_x (~ 1.12–$1.16\,\sigma_x$).

A similar relationship is expected for the colored-noise time series if the delay used in the reconstruction is greater than the correlation time. However, for a deterministic time series we expect this $(d+1)$ component's distance to be small for true NNs. The estimation results using our synthetic time series without removing temporal correlations are shown in Fig. 7(a), where we plot the mean value of this distance at each embedding dimension. As we can see, the analytical estimates are very close to the numerical results for the white random data sets, which are clearly larger than the ones for the deterministic time series.

For the correlated stochastic data we have an initial rapid drop in the mean $(d+1)$ components' distance followed by a slower gradual decrease. The initial drop is due to the deterministic correlations in the data caused by small delay and is more pronounced for weakly stochastic sets. The consequent slow decrease is due to the longer temporal correlations in the data, and it is more drastic for the strongly stochastic sets. All the correlated stochastic trends seem to level our at about 30 % of σ_x. The $(d+1)$-th distances for the deterministic data start near $0.5 \sim 0.6$ FNN fraction and drop down rapidly to near zero at the true minimal embedding dimension and then have slow gradual increase Fig. 7(a). This gradual increase is caused by the maximal local divergence rate of the nearby trajectories, and is expected to follow $e^{\lambda_{\max} d\tau_s}$ similar to Eq. (17).

---

[1] If $x_i \sim \mathcal{N}[0,\sigma_x]$, then $x_i - x_j \sim \mathcal{N}[0,2\sigma_x]$, and $\frac{|x_i - x_j|}{2\sigma_x} \sim \chi_1$.

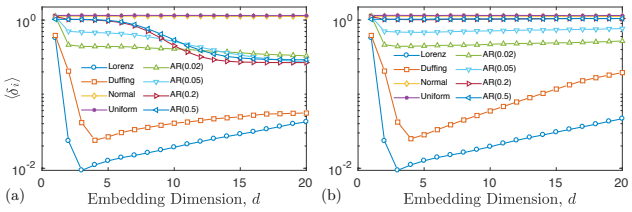

Fig. 7. Expected value of distances between the $(d+1)$-th coordinates of the NNs in d dimensions without (a) and with (b) temporarily decorrelated NNs

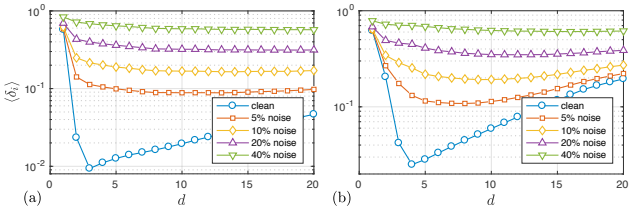

Fig. 8. Expected NN distance $\langle \delta_i \rangle$ versus the embedding dimension d for noisy Lorenz (a) and Duffing (b) time-series

The same plots were regenerated for the temporarily decorrelated nearest-neighbor statistics as shown in Fig. 7(b). The decorrelation had big effect on the correlated stochastic data: they were pushed up to the white random trends for the correlated data with larger stochastic components. In contrast, the AR(0.02) and AR(0.05) cases show trends more similar to the deterministic data, by exhibiting initial drops followed by the very slow increase. All the stochastic data stays above $0.45\sigma_x$ level and can be clearly differentiated from the deterministic data. The deterministic trends did not change during the initial drop to zero. However, the following gradual increase has a more pronounced character, especially for the Duffing data. This increase is more in line of expected exponential divergence caused by the maximal Lyapunov exponent, and was not as pronounced before due to temporarily correlated components staying close irrespective of embedding dimension.

The statistical analysis described in this section can be summarized as:

1. Temporal correlations significantly alter the observed results and need to be accounted for in accurate FNN analysis (removal of temporarily correlated points from the FNN analysis increases the differentiating power of the NN statistics).
2. Of the considered temporarily decorrelated metrics, $(d+1)$-th coordinate distance (δ_i)—as shown in shown in Fig. 7(b)—provides the best indication of the minimal embedding dimension and superior separation between the deterministic and random/stochastic trends, which remain constant with the increase in d. Thus, this metric can be used directly for estimating the embedding dimension for the noise free deterministic data. However, for noise contaminated deterministic data, it loses its ability to clearly identify the minimal embedding dimension as shown in Fig. 8.

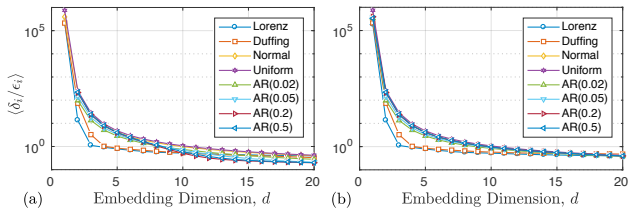

Fig. 9. Expected value of the Kennel ratio in d dimensions without (a) and with (b) temporarily decorrelated NNs

3. Both $\langle \varepsilon_i \rangle$ and $\langle \delta_i \rangle$ for deterministic data show exponential growth after minimal embedding dimension is reached. Thus, the original FNN fraction shown in Fig. 9 for both without (a) and with (b) temporarily decorrelated NNs is a good differentiator of deterministic data for low embedding dimensions. Setting a threshold value near $5 \sim 10$, this ratio with temporal decorrelation provides accurate estimates of FNN for $d < 10$, and clear separation of deterministic trends.

5 New Composite-Fraction-Based Metrics

Using the results of the statistical analysis described in the previous section, and combining them with the original definition of FNN, the following FNN *composite fraction* can be advocated:

$$R_{nn}(d;r,s) \triangleq \text{Prob}\left\{ (\delta_i > r\varepsilon_i) \vee (\delta_i > s\sigma_x) \right\}. \tag{20}$$

This new ratio has two parameters: r is the same as for the original fraction, while s is the same as for the reliable fraction. The first part is expected to serve as good indication of FNN for low d, while the second provides good indication of minimal embedding dimension and best separation of deterministic data from random/stochastic for larger d.

This composite metric still does not address the problem of NNs being close to each other due to the presence of noise instead of either dynamics or projection. Therefore, we propose a procedure similar to the FNS, where instead of looking up just one NN (NN) for each point $\mathbf{x}_i$, we look up k NNs. Then, the true, temporarily uncorrelated NN (i.e. the one close only due to dynamics or projection) is identified by evaluating the preceding and future l states of all k NNs. In particular, we evaluate the average distances between the base strand and all k NN strands of length $2l+1$. The strand with the smallest average distance from the base strand will contain the needed true NN.

5.1 New Composite-Fraction-FNN Algorithm

The composite FNN procedure can be summarized as follows:

1. For each point $\mathbf{x}_i$, identify its temporarily uncorrelated k NNs $\{\mathbf{x}_{j(m)}\}_{m=1}^{k} = \{\mathbf{x}_{j(m)} : |i-j(m)| > w,\ m = 1,\ldots,k\}$ just like in Eq. (13) to eliminate temporarily correlated NNs.
2. Then the true NN $\mathbf{x}_j$ is identified as:

$$j(m) = \arg\min_{m} \frac{\sum_{q=-l}^{l} \left\|\mathbf{x}_{j(m)+q} - \mathbf{x}_{i+q}\right\|}{(2l+1)\left\|\mathbf{x}_{j(m)} - \mathbf{x}_i\right\|}, \tag{21}$$

and used in Eq. (20) to determine FNN fraction.

Equation (21) indicates that the composite fraction will also be a function of l and k parameters:

$$R_{nn} = R_{nn}(d; r, s, l, k)\,. \tag{22}$$

It is reasonable to set the value of $l \leq w/2$ in Eq. (21) since we are focusing on temporal correlations. The choice of k can be influenced by many factors, such as noise level, density of points in the embedding, and sampling time. Here we only present results obtained with just one NN point ($k = 1$ and $l = 0$) for noise free data. In addition, we use $k = 10$ and $l = \tau$ for data with additive noise. However, we explicitly set these values in what follows and do not study their influence on the quality of the results, which is left for future work. Therefore, $R_{nn} = R_{nn}(d; r, s, \tau, k) \triangleq R_{nn}(d; r, s)$ in what follows.

The results using Eq. (20) are shown in Fig. 10 for fixed $s = 0.5$ (left plots) and for fixed $r = 4$ (right plots) for all the test data. For the deterministic data, the algorithm works well for $r \geq 4$ and for all $s > 0.2$ values, while for the same r values it provides correct indication of high-dimensional data for the white and correlated stochastic time series. In particular, the actual results are indistinguishable from the surrogate analysis for the random data and the composite fraction never drops below 50 %. For correlated stochastic data with a smaller stochastic component, we can clearly see that there is some deterministic component to the data, while for a large stochastic component, the results are very similar to the ones for the white noise. However, there still is the small but clearly identifiable separation between the surrogate and correlated stochastic data. This is especially evident for $s = 0.5$ plits, where for larger values of d there is a clear convergence in the FNN composite fraction to the same constant value for the surrogates, and to different values for all the correlated stochastic data sets. The separation of this surrogate FNN fraction from the actual FNN fraction seems to scale with d and the relative magnitude of the deterministic component. Therefore, we can use this separation between the actual and surrogate data composite fractions as an indication of some deterministic trends in the stochastic data. However, this needs further investigation for practical applicability in the analysis, with the purpose of relating the observed differences with the surrogates to the relative strength of the deterministic components.

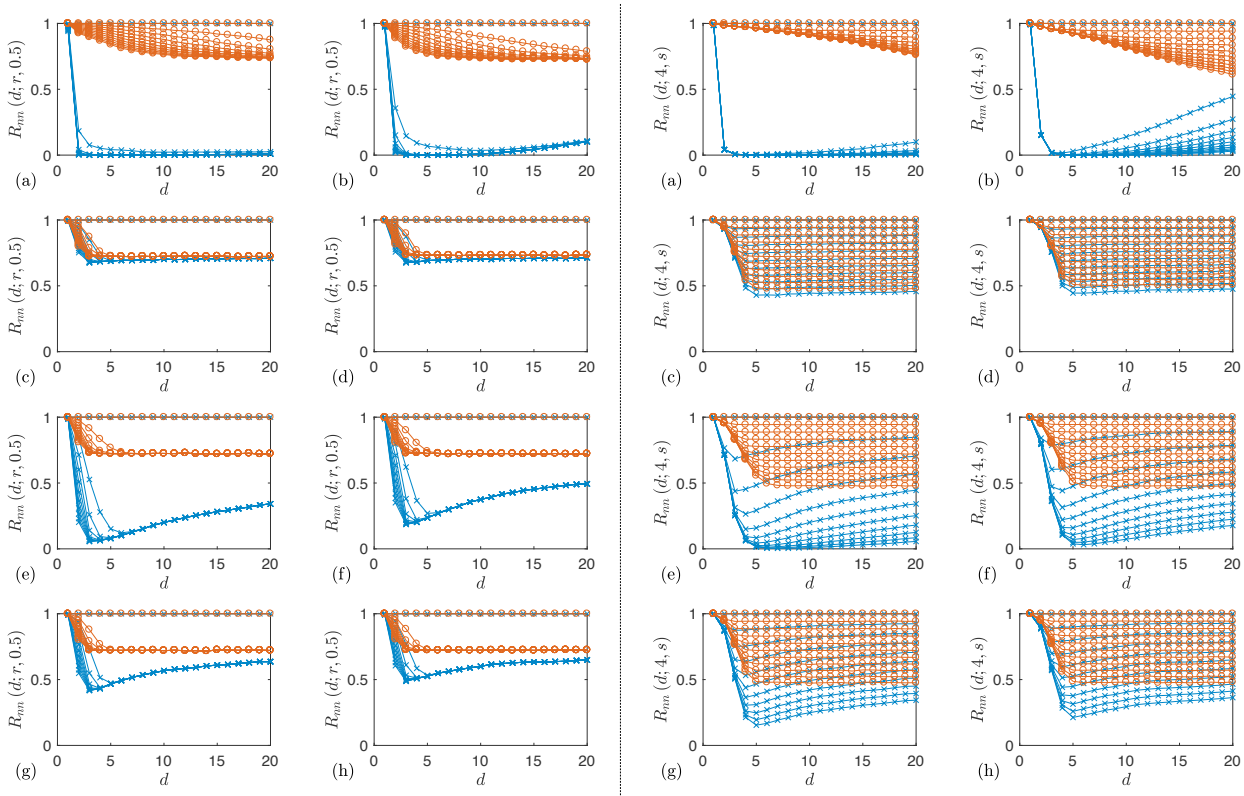

Fig. 10. Eq. (20) based composite fraction FNN algorithm with $s = 0.5$ (left plots) and with $r = 4$ (right plots) for Lorenz (a), Duffing (b), Normal (c), Uniform (d), AR(0.02) (e), AR(0.05) (f), AR(0.2) (g), and AR(0.5) (h). The gradual decrease in the fractions is precipitated by incrementing r from zero to 20 by two and s from zero to two by 0.1. Lines with circles reflect the corresponding surrogate data results.

One may expect the removal of temporal correlations and the new definition of NN points to positively alter the Kennel fraction results. However, the observed improvements were marginal: (1) marginally better indication of minimal embedding dimension for the deterministic data; (2) removal of artificial separation between the surrogates and actual data for the stochastic time series; and (3) accentuation exponential divergence for the deterministic data for larger embedding dimensions. Thus, the main deficiencies that were identified previously still remain.

5.2 New-Composite-Fraction FNS Algorithm

We have also modified the FNS algorithm to use a new metric specified by Eq. (20). We use exactly the same procedure as before, except instead of Eq. (15) we use:

$$R_{ns}(d;r,s) \triangleq \text{Prob}\left\{ \langle \delta_i \rangle_k > r \langle \varepsilon_i \rangle_k \vee \langle \delta_i \rangle_k > s\sigma_x \right\}, \tag{23}$$

where

$$\langle \varepsilon_i \rangle_k \triangleq \langle \varepsilon_i \rangle_{i \in S_k} = \frac{1}{\|S_k\|} \sum_{i \in S_k} \varepsilon_i \,. \tag{24}$$

The results of calculation using Eq. (23) are shown in Fig. 11 for fixed $s = 0.5$ (left plots) and for fixed $r = 10$ (right plots) for all the test data. For all the simulations we have used $w = \tau$. The differences with the new FNN are similar to what was observed previously in the reliable algorithm. The decrease of the FNS composite fraction to zero is slower and more gradual than for the FNN composite fraction for the deterministic data. However, the FNS composite fraction shows a much clearer indication for $d = 3$ for the Lorenz and $d = 4$ for the Duffing data as opposed to FNN, where we have to plot data on a log scale to clearly see the minima at the same values. This algorithm can be clearly used to differentiate the deterministic versus stochastic or random data using surrogate analysis. In addition, for the smaller stochastic component data, the surrogate analysis also provides the needed indication of the deterministic component. However, this indication is absent for the larger stochastic components for which they are indistinguishable from the random data sets.

6 Noise Robustness of Composite Fraction Algorithms

To test the applicability of these new algorithms for the noisy deterministic time series we have tested them on both Lorenz and Duffing data contaminated by additive noise as described before. The composite FNN fraction results for noisy Lorenz and Duffing are shown in Fig. 12 for $k = 1$ NN (left plots) and for $k = 10$ NNs (right plots), respectively. For both cases, we set $r = 10$ and $s = 1$. Plots demonstrate that using ten NNs in the algorithm considerably improves on results using just one NN. Composite FNN fractions with ten NNs have a clear inflection point, or elbow, in their trends at the appropriate minimal embedding dimension for low noise levels. Even for high noise levels the estimates for Duffing are good and for Lorenz only overestimate the embedding dimension by one.

New composite FNS plots for noisy Lorenz and Duffing data are shown in Fig. 13 left and right plots, respectively. Compared to the composite FNN fraction plots, these do not show as clear of an indication when the FNS composite fraction levels out or has a clear inflection point at the appropriate minimal embedding dimensions. Instead, it has a more gradual transition requiring some careful thresholding to identify the appropriate minimal embedding dimensions. It should be noted that these results are similar to the ones obtained for the stochastic time series with small stochastic components. Therefore, we have to be careful when making any judgments about the nature of additive noise versus systematic stochasticity. However, we can still clearly use these new methods in both cases to identify the deterministic components in the data and their dimensionality.

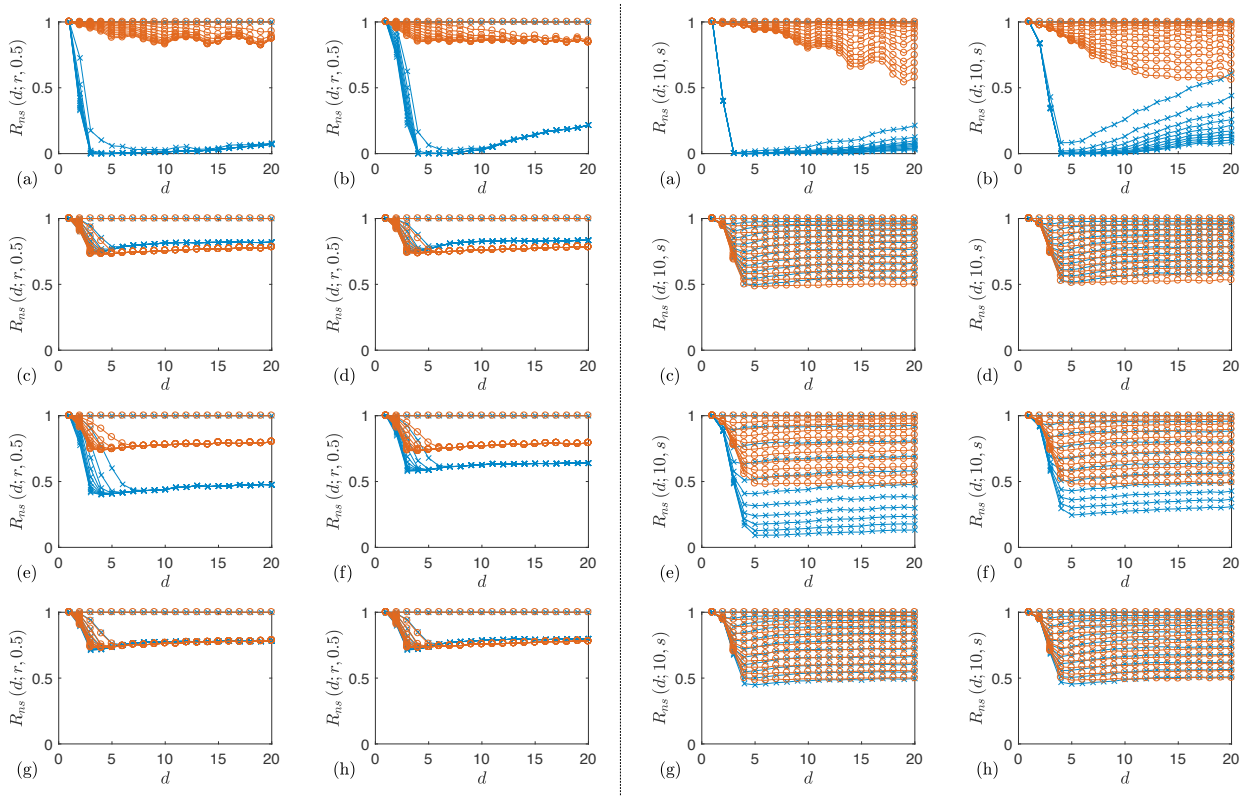

Fig. 11. Eq. (23) based composite fraction FNS algorithm with $s = 0.5$ (left plots) and with $r = 10$ (right plots) applied to Lorenz (a), Duffing (b), Normal (c), Uniform (d), AR(0.02) (e), AR(0.05) (f), AR(0.2) (g), and AR(0.5) (h). The gradual decrease in the fractions is precipitated by incrementing s from zero to two by 0.1 and r from zero to 20 by two. Lines with circles reflect the corresponding surrogate data results.

7 Summary and Conclusions

We have described currently used false nearest neighbors (FNN) algorithms, and have showed their shortcomings using synthetic deterministic, white random, and correlated stochastic time series. The effect of additive noise on the performance of the algorithms for the deterministic data was also identified and discussed. To understand the reasons behind the observed deficiencies, we have studied the statistical properties of the nearest-neighbor metrics used in the algorithms for all data types. It was observed that the expected $(d+1)$-dimensional distance between the d-dimensional nearest neighbors was a good indicator of minimal embedding dimension for the clean deterministic time series, but it lost its differentiation when even small additive noise was introduced into the deterministic data.

The results of the nearest neighbor analysis were used to derive new algorithms based on the *composite fraction* that overcome the deficiencies of the earlier methods, and are applicable even for noisy deterministic data. The concepts of FNN and false nearest strands (FNS) were also contrasted in this new composite framework. In addition, the FNN framework

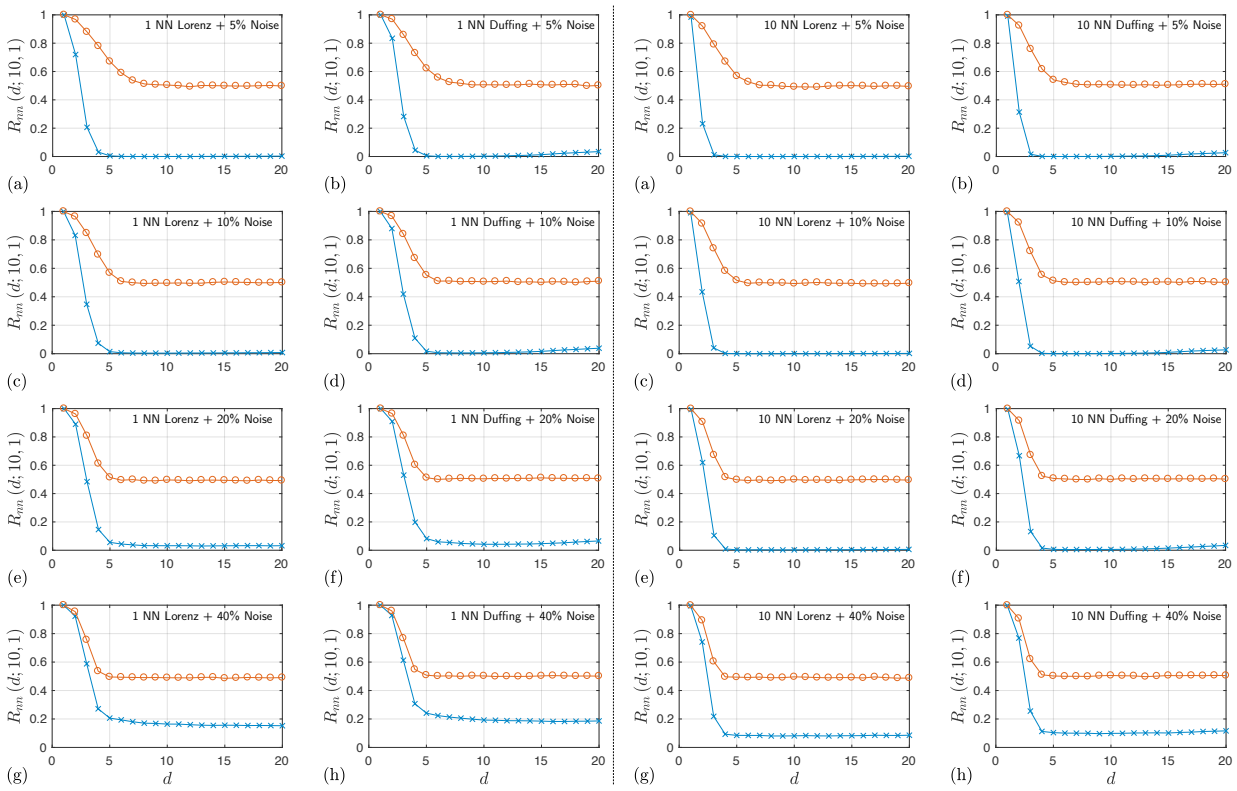

Fig. 12. Eq. (22) based FNN algorithms with $k = 1$ (left plots) and $k = 10$ (right plots) for Lorenz (first and third columns) and Duffing (second and fourth columns) and the surrogate data with 5 % (a,b), 10 % (c,d), 20 % (e,f), 40 % (g,h) additive noise levels. Lines with circles reflect the corresponding surrogate data results.

was enhanced by the ability to identify points that are nearest neighbors only due to dynamics or projections by estimating the minimal distance to the several nearest strands in d-dimensions, which considerably improved the results for the noisy data. While both concepts provide results that can be used for estimating the minimal embedding dimension even in a noisy data set, the composite FNS algorithm provides a clearer indication of the minimal embedding dimension for the noise free data compared to the composite FNN. However, the FNN algorithm has a more differentiating power and provides less ambiguous results for the deterministic data contaminated with additive noise. These FNN/FNS composite fractions are also suitable in evaluating the presence and relative magnitude of a deterministic component in a stochastic or noisy data set.

Acknowledgments

This paper is based upon work supported by the National Science Foundation under Grant No. 1100031.

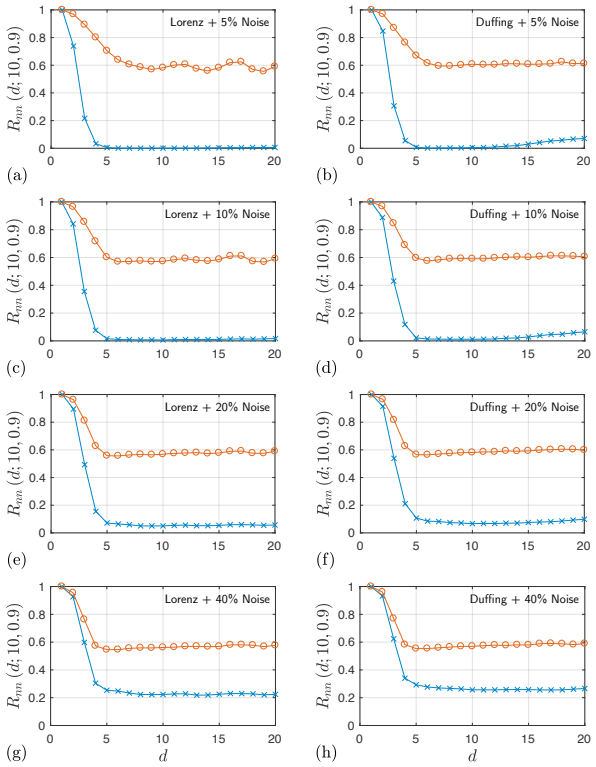

Fig. 13. Eq. (22) based FNS algorithms for Lorenz (left column) and Duffing (right column) and the surrogate data with 5 % (a,b), 10 % (c,d), 20 % (e,f), 40 % (g,h) additive noise levels. Lines with circles reflect the corresponding surrogate data results.

References

[1] Takens, F., 1981. "Detecting strange attractors in turbulence". In *Dynamical systems and turbulence, Warwick 1980*, D. A. Rand and L. S. Young, eds. Springer-Verlag, Berlin, pp. 366–381.

[2] Sauer, T., Yorke, J. A., and Casdagli, M., 1991. "Embedology". *Journal of statistical Physics,* **65**(3-4), pp. 579–616.

[3] Molkov, Y. I., Mukhin, D., Loskutov, E., Feigin, A., and Fidelin, G., 2009. "Using the minimum description length principle for global reconstruction of dynamic systems from noisy time series". *Physical Review E,* **80**(4), p. 046207.

[4] Pecora, L. M., Moniz, L., Nichols, J., and Carroll, T. L., 2007. "A unified approach to attractor reconstruction". *Chaos: An Interdisciplinary Journal of Nonlinear Science,* **17**(1), pp. 013110–013110.

[5] Kember, G., and Fowler, A., 1993. "A correlation function for choosing time delays in phase portrait reconstructions". *Physics Letters A,* **179**(2), pp. 72–80.

[6] Fraser, A. M., 1989. "Reconstructing attractors from scalar time series: a comparison of singular system and redundancy criteria". *Physica D: Nonlinear Phenomena,* **34**(3), pp. 391–404.

[7] Fraser, A. M., and Swinney, H. L., 1986. "Independent coordinates for strange attractors from mutual information".

Physical review A, **33**(2), p. 1134.

[8] Grassberger, P., Schreiber, T., and Schaffrath, C., 1991. "Nonlinear time sequence analysis". *International Journal of Bifurcation and Chaos,* **1**(03), pp. 521–547.

[9] Broomhead, D. S., and King, G. P., 1986. "Extracting qualitative dynamics from experimental data". *Physica D: Nonlinear Phenomena,* **20**(2), pp. 217–236.

[10] Gibson, J. F., Doyne Farmer, J., Casdagli, M., and Eubank, S., 1992. "An analytic approach to practical state space reconstruction". *Physica D: Nonlinear Phenomena,* **57**(1), pp. 1–30.

[11] Rosenstein, M. T., Collins, J. J., and De Luca, C. J., 1994. "Reconstruction expansion as a geometry-based framework for choosing proper delay times". *Physica D: Nonlinear Phenomena,* **73**(1), pp. 82–98.

[12] Kugiumtzis, D., 1996. "State space reconstruction parameters in the analysis of chaotic time seriesthe role of the time window length". *Physica D: Nonlinear Phenomena,* **95**(1), pp. 13–28.

[13] Cao, L., Mees, A., and Judd, K., 1998. "Dynamics from multivariate time series". *Physica D: Nonlinear Phenomena,* **121**(1), pp. 75–88.

[14] Judd, K., and Mees, A., 1998. "Embedding as a modeling problem". *Physica D: Nonlinear Phenomena,* **120**(3), pp. 273–286.

[15] Kim, H., Eykholt, R., and Salas, J., 1999. "Nonlinear dynamics, delay times, and embedding windows". *Physica D: Nonlinear Phenomena,* **127**(1), pp. 48–60.

[16] Small, M., and Tse, C. K., 2004. "Optimal embedding parameters: a modelling paradigm". *Physica D: Nonlinear Phenomena,* **194**(3), pp. 283–296.

[17] Maus, A., and Sprott, J., 2011. "Neural network method for determining embedding dimension of a time series". *Communications in Nonlinear Science and Numerical Simulation,* **16**(8), pp. 3294–3302.

[18] Chatzinakos, C., and Tsouros, C., 2015. "Estimation of the dimension of chaotic dynamical systems using neural networks and robust location estimate". *Simulation Modelling Practice and Theory,* **51**, pp. 149–156.

[19] Kantz, H., and Schreiber, T., 2004. *Nonlinear time series analysis*, Vol. 7. Cambridge university press.

[20] Abarbanel, H., 1996. *Analysis of observed chaotic data*. Springer, New York.

[21] Kennel, M. B., and Abarbanel, H. D., 2002. "False neighbors and false strands: A reliable minimum embedding dimension algorithm". *Physical review E,* **66**(2), p. 026209.

[22] Kennel, M. B., Brown, R., and Abarbanel, H. D., 1992. "Determining embedding dimension for phase-space reconstruction using a geometrical construction". *Physical review A,* **45**(6), p. 3403.

[23] Cao, L., 1997. "Practical method for determining the minimum embedding dimension of a scalar time series". *Physica D: Nonlinear Phenomena,* **110**(1), pp. 43–50.

[24] Hegger, R., and Kantz, H., 1999. "Improved false nearest neighbor method to detect determinism in time series data". *Physical Review E,* **60**(4), p. 4970.

[25] Kazem, A., Sharifi, E., Hussain, F. K., Saberi, M., and Hussain, O. K., 2013. "Support vector regression with chaos-based firefly algorithm for stock market price forecasting". *Applied Soft Computing,* **13**(2), pp. 947–958.

[26] Hanslmeier, A., Brajša, R., Čalogović, J., Vršnak, B., Ruždjak, D., Steinhilber, F., MacLeod, C., Ivezić, Ž., and Skokić, I., 2013. "The chaotic solar cycle-ii. analysis of cosmogenic 10be data". *Astronomy & Astrophysics,* **550**, p. A6.

[27] Wernitz, B., and Hoffmann, N., 2012. "Recurrence analysis and phase space reconstruction of irregular vibration in friction brakes: Signatures of chaos in steady sliding". *Journal of Sound and Vibration,* **331**(16), pp. 3887–3896.

[28] Hampson, K. M., and Mallen, E. A., 2012. "Chaos in ocular aberration dynamics of the human eye". *Biomedical optics express,* **3**(5), pp. 863–877.

[29] Lorenz, E. N., 1963. "Deterministic nonperiodic flow". *Journal of the atmospheric sciences,* **20**(2), pp. 130–141.

[30] Duffing, G., 1918. *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung*. No. 41-42. R, Vieweg & Sohn.

[31] Sprott, J. C., and Sprott, J. C., 2003. *Chaos and time-series analysis*, Vol. 69. Oxford University Press Oxford.

[32] Schreiber, T., and Schmitz, A., 1996. "Improved surrogate data for nonlinearity tests". *Physical Review Letters,* **77**(4), p. 635.

[33] Hegger, R., Kantz, H., and Schreiber, T., 1999. "Practical implementation of nonlinear time series methods: The tisean package". *Chaos: An Interdisciplinary Journal of Nonlinear Science,* **9**(2), pp. 413–435.

[34] Golub, G. H., and Van Loan, C. F., 2012. *Matrix computations*, Vol. 3. JHU Press.

[35] Theiler, J., 1991. "Some comments on the correlation dimension of $1/f^{\alpha}$ noise". *Physics Letters A,* **155**(8), pp. 480–493.

[36] Theiler, J., 1990. "Statistical precision of dimension estimators". *Physical Review A,* **41**(6), p. 3038.

[37] Gut, A., 2009. *An intermediate course in probability*. Springer.

List of Figures